

In the Claims:

1-118. (Canceled)

119. (Currently amended) A method for automating the extraction of information from a semi-structured document characterized by a document type that comprises design and structural characteristics of a set of similar documents, the method comprising:

designing a target extraction template for the terms of the document type;

supporting the creation of a control set of documents containing the terms manually tagged to the extraction template;

automatically generating a skeleton of an extraction model tree for every term;

identifying a set of selectors for each model tree;

training the models trees by automatically ~~optimizing~~ identifying a subset of the selectors for the ~~of the term~~ extraction models trees for ~~to the best~~ compliance with the control set ~~tagging;~~ and

extracting information from the document with the optimized model trees using the ~~optimized model to automatically extract information from the document; and~~

storing the extracted information in a database.

120. (Previously presented) The method of claim 119, further comprising using specialized invariants to select generic components of information from the document.

121. (Previously presented) The method of claim 119, further comprising tracking and analyzing changes made to initially extracted information and subsequent re-optimization of models.

122. (Currently amended) The method of claim 119, further comprising analyzing an additional semi-structured document and updating the ~~model~~ selectors or its structure if a change

in accuracy of the term extraction model exceeds a threshold.

123. (Currently amended) The method of claim 119, further comprising:

(a) retaining specific information about a set of semi-structured documents to serve as a template for new semi-structured document introduction;

(b) comparing any new semi-structured document with a pattern represented by specific information known to be suitable for searching for text based on the retained specific information about the set of semi-structured;

(c) assessing if the ~~result~~ comparison of (b) is within a threshold of the result of (a).

124. (Previously presented) The method of claim 123, as applied to knowledge that a given company employs similar patterns for subsequent versions of similar documents identifying the company to which the document pertains.

125. (Previously presented) The method of claim 119, in which terms can be assigned a term class for at least one of immediate validation, synonym support, and vocabulary management.

126. (Previously presented) The method of claim 119, further comprising automatically comparing first and second extracted data to each other to identify extraction errors.

127-136. (Withdrawn)

137. (New) A method, comprising:

identifying a plurality of indicators in a document type;

generating a decision tree for the document type based on a subset of the plurality of indicators;

identifying a location of a term within the document type as a function of the decision tree;

comparing the location of the term with a control location for the term in the document

type; and

generating an extraction template for the document type.

138. (New) The method according to claim 137, further comprising:

determining whether the location of the term within the document type is at least one of a title, a sentence, a narrative, an interrogative sentence, an exclamatory sentence, a paragraph and a table.